

Use of Public Transport Smart Card Fare Payment Data for Travel Behaviour Analysis in Singapore

Conference Paper**Author(s):**

Chakirov, Artem; Erath, Alexander

Publication date:

2011

Permanent link:

<https://doi.org/10.3929/ethz-a-006742061>

Rights / license:

In Copyright - Non-Commercial Use Permitted

Originally published in:

Arbeitsberichte Verkehrs- und Raumplanung 729

USE OF PUBLIC TRANSPORT SMART CARD FARE PAYMENT DATA FOR TRAVEL BEHAVIOUR ANALYSIS IN SINGAPORE

A. CHAKIROV ^a, A. ERATH ^a

^a *Future Cities Laboratory,*

Singapore -ETH Centre for Global Environmental Sustainability (SEC),

Singapore

Email: chakirov@ivt.baug.ethz.ch

ABSTRACT

Electronic payment systems for public transport, introduced in many cities all over the world during the last decade, are able to capture system transactions and provide comprehensive data records on usage of public transport. Processing and analysis of this data opens new opportunities in transportation and travel behaviour research and is becoming an emerging research topic.

This paper presents an analysis of one full day of public transport smart card activity for the entire city-state of Singapore and investigates its potential for characterization of public transport system and urban travel behaviour. In particular, an assessment of spatial and temporal travel behaviour including mode choice, travel- and waiting times is performed using statistical tools and data mining techniques. Thereby data reveals country specific preferences and behaviours as the influence of high passenger volumes and seat availability on route choice decisions and travel times. Furthermore data's ability to detect travel patterns and infer peoples travel purposes and locations of regular activities is investigated and strategies for modelling of home and work locations are presented. Additionally, household survey data from 2008 is used for comparison of multi modal travel patterns involving public and private transport.

Keywords: Public transport, smart card data, travel behaviour

1. INTRODUCTION

Electronic fare collection as the most advanced form of automatic fare collection (AFC) systems for public transport offers many advantages and benefits for operators of public transport as well as their customers. Convenient, easy and almost instantaneous payment process saves customers time and makes use of public transport more attractive. Furthermore, lower operation costs, high efficiency and reliability as well as new opportunities for implementation of flexible fare schemes are additional benefits for operators.

Another important aspect of an electronic fare collection system is the detailed data records that are continuously generated and archived. Analysis of such data can provide valuable insights into usage of public transport and help for better understanding of people's travel behaviour and their preferences. Consequently, processing and mining of these substantial amounts of data are becoming more and more emerging research topics in the areas of mobility and transportation planning.

In this context, the example of Singapore represents an interesting study case as a highly dynamic city with a centralized, long-term oriented transportation planning policy and fast growing commuter population. The combination of a distance-based fare scheme for public transport and the use of contactless stored-value cards, so called EZ-Link cards, for payment of public transport fares, provide a nearly comprehensive data record of public transportation usage for the entire city. This allows detailed assessment of travel behaviour and mobility patterns, which is the topic of this work. The main goals thereby are the detailed detection and description of specific travel behaviour in Singapore and use of AFC data for activity location modelling.

This paper starts with a brief overview of recent research on analysis and use of smart card data. Afterwards, in section 3, the AFC system implemented in Singapore is introduced and the available data records are described. The following section 4 covers the analysis of travel behaviour in Singapore with investigation of specific preferences and usage patterns of public transport in particular. In section 5, the focus is switched to identification of activities, performed between different public transport journeys. Additional data needs and further future strategies for more comprehensive activity modelling are presented. The paper concludes with future outlook and a discussion of opportunities and limitations of the smart card data use for travel behaviour characterization and transport planning.

2. LITERATURE REVIEW

In recent years the number of published articles and papers on smart card data use in public transport is increasing. A comprehensive literature review on this topic is offered by Pelletier *et al.* (2011), where three categories of studies have been identified: strategic level studies with focus on long-term planning, tactical level studies related to schedule adjustment, load profiles and transfers and operational-level studies analysing performance indicators of public transport network and smart card operations itself. Research with special focus on spatial travel information provided by AFC systems has been done by Li *et al.* (2011) and Munizaga *et al.* (2011), where Origin-Destination Matrices are estimated using smart card data. Moreover recent work with focus on commercialization by Páez *et al.* (2011) presents a geodemographic analysis, which links smart card data to household survey data and business data points to identify potential commercial partnerships near metro stations in Montreal.

In addition work on use of smart card data very similar to Singapore's public transport AFC system was done in Seoul. Park *et al.* (2008) investigated reliability of smart card data and its potential for describing characteristics of public transport usage. Jang (2010) uses one week data from the same AFC system to analyse travel time and transfer patterns in Seoul.

However, only few research papers address modelling of activity types and locations implicitly based

on travel data. Ahas *et al.* (2006) used positioning data from mobile phones to identify activity locations meaningful to users. Chu *et al.* (2010) presented a methodology for characterization of trips based on socio-demographic characteristics, multiday travel patterns and association of travel with specific locations. Strategies for extension of these concepts towards more general activity location modelling based on smart card data is one of the goals of this work.

3. EZ-LINK SMART CARD IN SINGAPORE

3.1 Automatic face collection system

The AFC system for public transport based on contactless, stored value smart cards, so called EZ-Link cards, was introduced in Singapore in April 2002. Today, the EZ-Link card can be used island wide for payment of all modes of public transport, regardless of operator as well as for minor retail transactions, parking and road toll payment. Though cash payment of single fares at higher rates is still possible, e-payments with EZ-Link card account for 96% of all trips, which makes the data records from EZ-Link card highly comprehensive and the missing cash paying travellers negligible (Prakasam, 2008).

The implementation of uniform smart card AFC system allowed the introduction of a distance-based fare scheme for all modes of public transport in Singapore. The fare charge for each customer is based exactly on the travelled distance, transport mode and demographic attributes as there are prioritized rates for children, students and senior citizens. Customers have to tap their EZ-Link card on the reading device every time they enter and leave a train station or a bus. Thus, besides of the information on boarding time and location, the data collected from EZ-Link cards contains detailed records of alighting times and destination location for both bus stops as well as Mass Rapid Transit (MRT) and Light Rail Transit (LRT) stations. These attributes distinguish the Singaporean smart card data from those collected by the majority of other automatic fare collection systems and allow more detailed assessment of travel behaviour and mobility patterns. Furthermore, as the EZ-Link cards are easily rechargeable, people tend to continuously use one single EZ-Link card with a unique card ID for all their public transport journeys for substantial periods of time. As the technical setup of the system doesn't allow more than one person to travel on a single EZ-Link card, it can be assumed that each unique card ID represents one single person. This enables highly disaggregated analysis of individual itineraries and opens new ways for understanding peoples travel behaviour on short as well as long term scales.

3.2 Available EZ-Link data set

In the context of this paper, we adopt definitions used by the Singapore Land Transportation Authority (LTA) and EZ-Link Corporation and define a journey as a one way travel from one activity to another. Each journey consists out of one or several consecutive journey stages or trips on same or different modes (LTA, 2011).

The analysis presented in this paper is based on one full day record of journeys from 5am of Tuesday, 15th September 2010 till 2am of Wednesday, 16th September 2010. This data set was already pre-processed by LTA and single trips of each customer were aggregated to journeys, according to fare rules defined by Public Transport Council (2010). For further processing of this data, we used MySQL open source database in combination with programming languages R and Java. Furthermore all train stations and bus stops were geocoded using information provided by LTA.

The original aggregated one-day data set contains ca. 3.6 million journey records and ca. 1.8 million unique card IDs. In the process of preparation and reviewing of this data, about 84'000 journeys (2.4%) were removed from the dataset due to errors, missing values or illogical journey routes, with same origin and destination location recorded as a result of aggregation according to the fare rules. It is also important to note that as this data set doesn't contain service numbers of bus routes, vehicle

IDs or interchange locations, its use for detection of weak points of public transport network, like overcrowded buses is limited.

4. TRAVEL BEHAVIOUR

4.1 Characteristics of public transport usage in Singapore

A variety of methods can be applied for characterization and statistical analysis of public transport usage in urban areas. Description in spatial and temporal dimensions are often used by transportation planning authorities as indicators for performance and quality of service. The objective of this section is not to provide comprehensive statistical description of Singaporean public transport system but to show how the EZ-Link data can be specifically used as indicator of region and culture specific travel behaviour and identification of external factors influencing peoples travel decisions.

In this context analysis of times, when people decide to start their travel can provide a valuable insight. Figure 5.1 shows the temporal distribution of boarding times at the first stop or station of the journey for all public transport journeys during 24h. The sharp peaks in the morning and evening hours are eye-catching. Especially the evening peak shows a high number of journey starts concentrated within a 15-20 min period. The fact that many people start their journeys almost synchronously during morning as well as evening peak hours indicates little flexibility in the travel time choice for commuting, which is assumed to be related to commonly inflexible working hours in Singapore. Additionally it is worth noting the small peak around 2pm in the afternoon might result from lunch time travels, part-time workers and class hours in educational institutions.

The shown trip distribution also highlights the main advantages of analysis based on AFC data over the commonly used self-reported survey. High temporal resolution and high reliability of records allow aggregation on a minute basis and so the detection of sharp peaks, clearly visible in Figure 5.1. This peak wouldn't look that distinct and be more diluted in case data of data aggregated in larger time bins. Exactly this though is usually required for less precise and personally biased data from self-reported travel diaries. Using AFC data instead transportation planners and transportation authorities can obtain information and gain new insights, hidden from them beforehand.

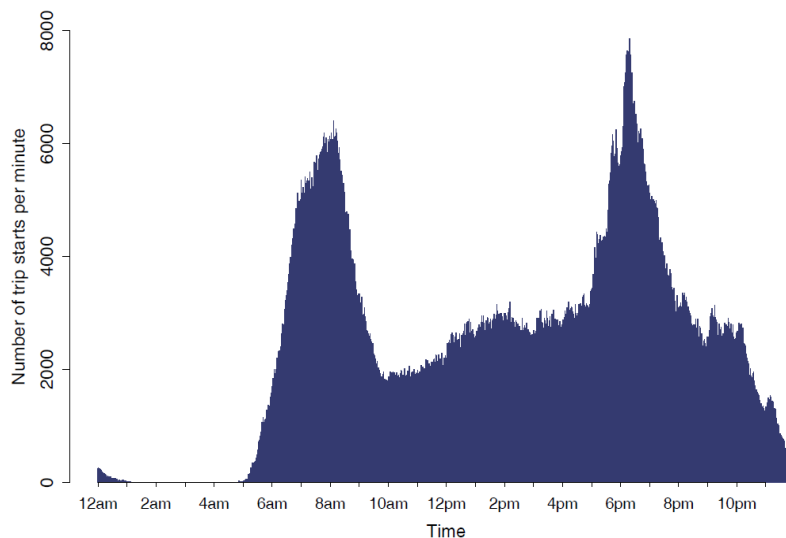


Figure 5.1. Temporal trip start distribution

4.2 Identification of Waiting Times

4.2.1 Identification Method

Another important temporal measure, which has implications on travel behaviour and route choice, is the waiting time – the timespan between arriving to the station or stop and actually boarding the transportation vehicle. In case of EZ-Link data, no information on waiting time of bus passengers can be obtained, as the smart card reading devices are located inside the buses and passengers tap their EZ-Link card as they enter the vehicle. However, in case of MRT and LRT, EZ-Link cards are tapped during entering or leaving the station and not the vehicle. This implies that the waiting time is a part of total recorded travel time and can be extracted from EZ-Link data record.

In this paper we use the term “waiting time” to describe the time, a passenger spends inside the station, which includes both walking and the actual waiting time. In order to extract this time for each station, all MRT trips from this station to other stations on the same MRT line are considered. Hence only direct, uninterrupted trips without interchanges are used for calculation of waiting times. For each pair of stations, travel times of passengers traveling between these two stations in each direction are extracted. It is assumed that the fastest passenger arrived to the platform just in time and did not have to wait for the train at all. His waiting time is considered to be zero and his travel time is used as a benchmark. By subtracting this benchmark time from the travel times of all other customers travelling between the same two stations, waiting times of these passengers are obtained.

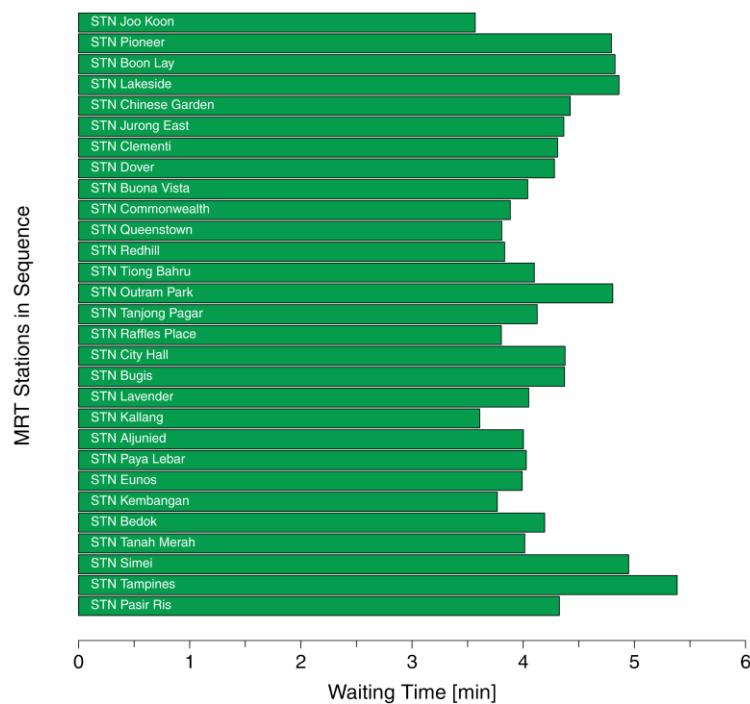


Figure 5.2. Average waiting times (in-station times) per station for East West MRT Line during 24h

4.2.2 Results and Findings

Dependent on the focus of the analysis of waiting times at MRT stations, information on significant service frequency changes during the day as well as potentially overcrowded trains can be gained. However, it is important to note, that due to the definition and identification method of the waiting time, the absolute time values don't precisely indicate the service frequency at the particular station. Large hub station with several access points tend to have larger variations and higher averages of waiting times, which reflects different lengths of access ways to the platform from different entry points. In Figure 5.2 the average waiting times per station for Singapore's East-West MRT line are shown. As expected large hub stations with long underground access ways like Outram Park Station

show longer waiting times. Notably eye-catching and the actual finding of such analysis is however the fact that the longest waiting time is found for Tampines – a single above-ground MRT station. This exceptionally high value in comparison to the other stations can be explained by looking at the histogram of waiting times, shown in Figure 5.3. Next to the common waiting time distribution around 4 min, a second peak of waiting times ranging from 10 to 15 min is clearly visible. This peak results from passengers, who take a train in opposite direction towards the last stop at Pasir Ris and then stay in the train to secure a seat for their journey during the morning peak hour. This observation leads to an interesting conclusions: the value of having a seat is exceptionally high and for some people it is worth about 10 min of additional travel time. This findings are also an indication for high passenger volumes and long MRT journeys. As in case of journeys from Tampines station, the average distance to destination station for passengers taking the detour over the Pasir Ris Station is 18.2 km compared to average 14.7 km for passengers taking the direct way.

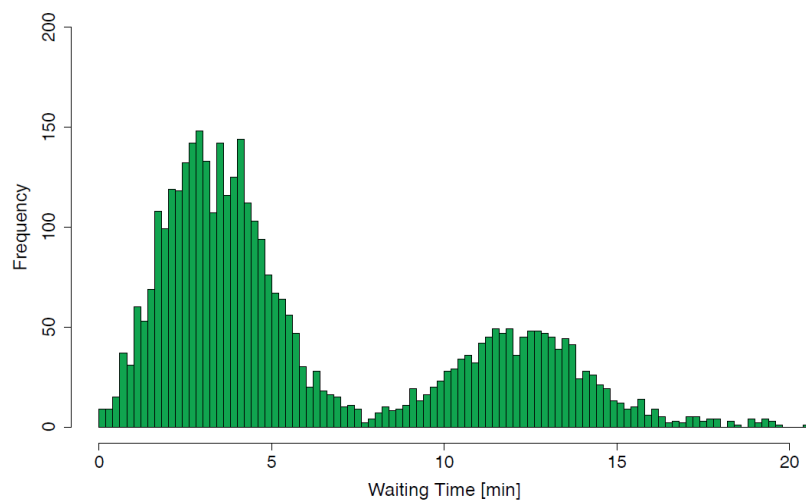


Figure 5.3. Histogram of waiting times at the Tampines MRT Station for journeys starting between 7am-9am

5. TOWARDS ACTIVITY LOCATION MODELLING BASED ON PUBLIC TRANSPORT JOURNEYS

5.1 Activities as reason for travelling

People usually travel from one place to another because they want to perform an activity, which cannot be performed in the desired way at their current location. The activities can be of various types and durations as there are work, education, leisure and social activities for example. Hence, in order to understand peoples travel behaviour and therewith flows, mobility patterns and traffic volumes, it is necessary to look at people's daily activities and their locations. This approach is also used in modern agent-based transport simulations like MATSim, which include spatial and temporal information about activities in its model (MATSim-T, 2011). In absence of comprehensive data sources such as a whole population census, one of the main challenges in modelling activities results from lack of verified data of high spatial resolution on people's home and particularly work locations. However such information is not only important for transport planning and implementation of agent-based transport models, but is also very valuable in the areas of urban development and land-use planning.. Travel patterns observed in smart card data from public transport can provide valuable information on peoples primary activity locations and help to verify and refine existing models and assumptions.

5.2 Consistency of public transport journeys chains

To make statements about activities of a particular person, the recorded daily journey chain of this person needs to be consistent. Consistency in the context of AFC smart card data record means, that

the person who arrived to the activity location by public transport, has to leave it after ending the activity also by public transport, otherwise the activity duration can't be extracted. The assumption of consistency is hard to verify based on AFC data only, as the use of any other means of transport except from public transport, e.g. walking, cannot be detected. However, the most obvious cases of inconsistency can be identified by analysing the distances between the alighting location of the last journey and the boarding location of the following journey. Using this method in-between trips by other means of transport such as taxi or car can be recognized.

The analysis of distances between EZ-Link journeys is shown in form of a cumulative relative frequency graph in Figure 6.1. Only persons with more than one journey recorded in the used one day EZ-Link record were evaluated. The graph shows that 89.7% of journeys, which are following a preceding one, start less than 1 km away from previous alighting location. This indicates that the majority of public transport users don't switch to other transport modes between public transport journeys.

A slightly different picture is obtained by investigating distances between first boarding and last alighting station of the day, which in case that they are close by, gives a strong indication for home location. As shown in the cumulative relative frequency graph of these distances for all persons in the EZ-Link data, including those with one journey (Figure 6.1), only 61% return to the station, which is located in radius of 1 km around their first departure station of the day. This is for the major part a logical consequence of the fact, that 30% of EZ-Link journeys are one-way journeys. In order to better understand the mode choice and identify transportation modes used in combination with public transport in Singapore, we make use of the Household Interview Travel Survey 2008 (Land Transportation Authority, 2009) in the following section.

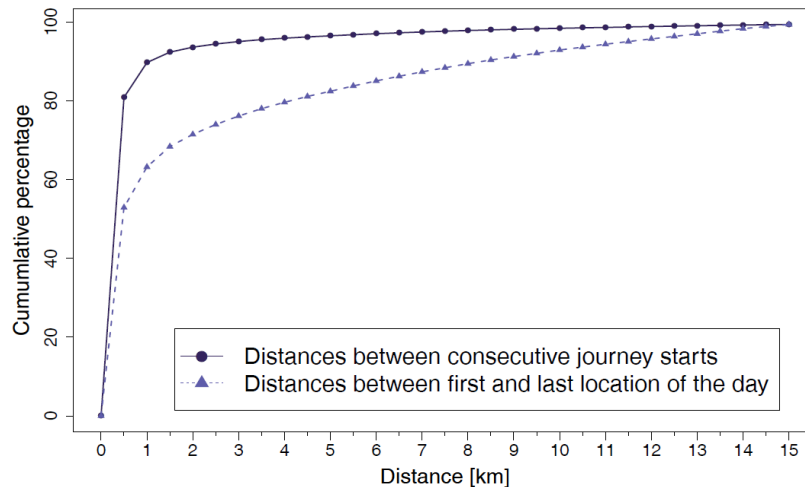


Figure 6.1. Cumulative relative frequency graph of distances between journey stages and distances between first boarding location and last alighting location of the day

5.3 Comparison with the Household Interview Travel Survey 2008

The Household Interview Travel Survey (HITS) is conducted by Land Transportation Authority in Singapore every 4 years and involved over 10500 households during its latest conduct in 2008 (Land Transportation Authority, 2009). The detailed travel diary data from this survey can be used as a valuable supplement for enrichment of the information obtained from analysis of the AFC smart card data.

In order to investigate the questions raised in previous section and better understand the mode-mix in Singapore, the daily journey record of persons using public transport at least once during the day, which is total of 13904 persons, is extracted from HITS 2008 and referred to as PT-HITS in following.

Looking at the distances between first starting point and last destination of the day, we observe more than 98% of all persons returning by the end of the day to the same locations. This result is expected as participants in HITS supposed to report all trips including all modes. Analysis of the alternative modes of transport used by PT-HITS interviewees beside public transport, can partially explain the gap, which we observed in Figure 6.1. A total of 17.4% of reported journeys in PT-HITS include or are completely conducted by modes different from public transport. We group these modes in 7 categories: passenger, private bus, taxi, driver, walk, which is considered a mode only if walking distance is longer than 1 km, cycle and an unknown mode. Trips with unknown or other mode is partially reported in HITS and partially result of identified underreporting, as in the case when the new journey starts more than 1 km away from previous alighting location. Figure 6.2 shows the distribution of these modes with the dark coloured share representing the percentage of first or last journeys of the day using this mode. The passenger mode, which reflects pick-ups and drop-offs by a family member or a friend driving a private vehicle or a motorcycle, is with 7.4% of all journeys of the day, the most popular one. This noticeable aspect of Singaporean travel behaviour is a result of a strong tendency to avoid the use of public transport for the first or last mile of the trip, combined with high costs of owning a car, which encourages car pools. Furthermore, private buses such as company, school or other shuttle buses are used for 5.1% all journeys in PT-HITS. Almost all of these journeys are made as a part of the first or last journey of the day, which indicates the high usage of private buses for travels to and from working or education location.

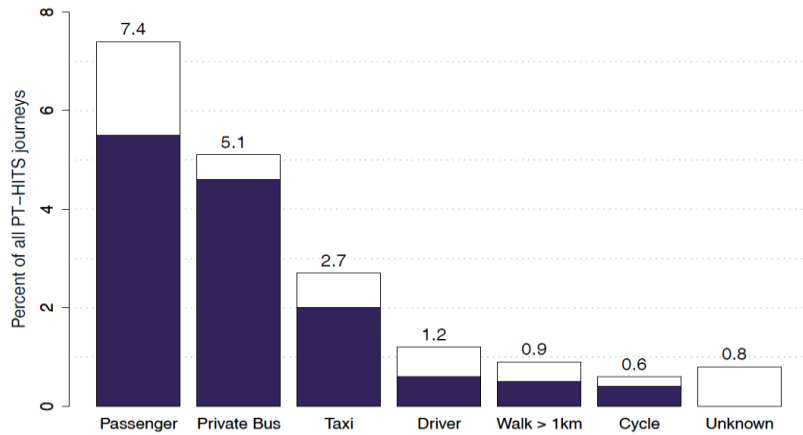


Figure 6.2. Distribution of additional transportation modes of public transport customers reported in PT-HITS. Dark coloured is the percentage of first or last journey of the day.

5.4 Activity durations and detection of work activities and location

One of the most important indicators for type of activity performed between two consecutive journeys is its duration. Figure 6.3 shows the distribution of activity durations between journeys recorded in EZ-Link data. As activity duration, we consider the time span between two journeys of a consistent journey chain as defined in section 6.1. Consequently again only persons with more than one journey per day and distances between consecutive journeys with less than 1 km are considered. The observed peak between approx. 8 and 12 hours indicates the typical working time span and can be used to identify working activities. The sharp edge at 45 min. is a result of aggregation method and definition of a journey by Public Transport Council (2010). According to these rules, 45 min is a time limit, under which two consecutive trips are considered as a part of one journey. This makes many of short term activities invisible and shows the overestimation of linked trips in the aggregated EZ-Link data. Observed activities under 45 min are mostly results of two journeys with MRT or LRT, as here the 45 min limit do not apply and each entrance to the station is considered as a new journey.

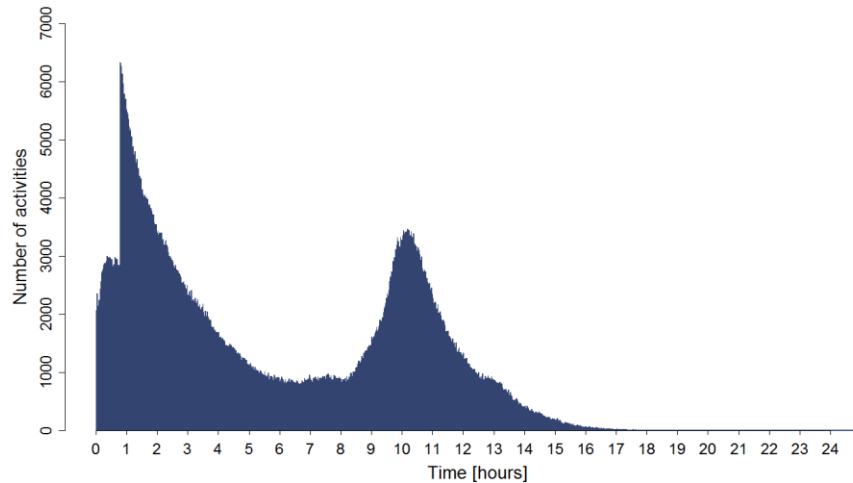


Figure 6.3. Distribution of activity durations

In the next step, the first estimation of work locations in Singapore based only on activity durations extracted from EZ-Link data was performed. It is assumed that all activities with duration between 8 and 12 hours are work activities. Figure 6.4 shows the geographical distribution of detected work location in Singapore aggregated according to 55 areas defined in Development Guide Plan by Urban Redevelopment Authority in Singapore. Only activities from consistent public transport journeys are considered, so work locations of people arriving to work with private transport as car, taxi or private busses are not included. To account for those in the final model, again data from HITS 2008 can be used.

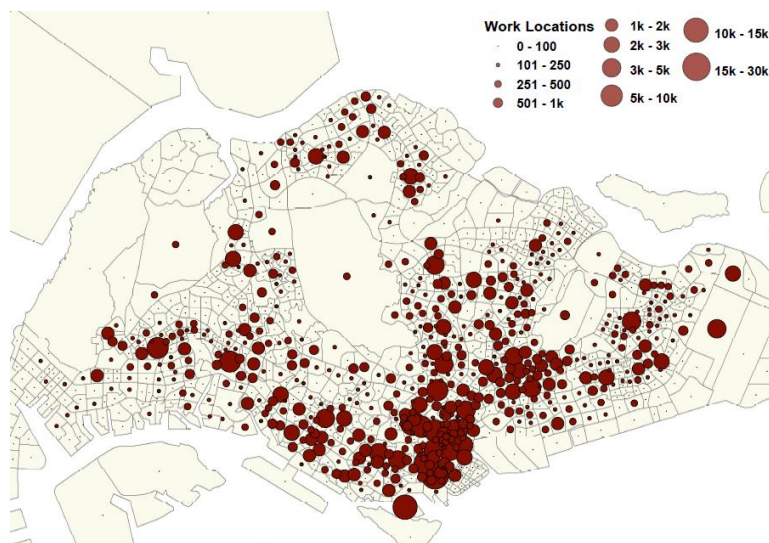


Figure 6.4. Distribution of detected work locations in Singapore

5.5 Further identification strategies for activity types and locations

The analysis of activity durations is a first step towards characterization of primary activities. In order to make this characterization more reliable and enable identification of home and work locations by use of EZ-Link data, further information is required.

One valuable source of additional information, as already shown in section 6.3 is the HITS data. As for instance country specific durations of certain activities can be extracted from HITS and used for more precise characterization of activities extracted from EZ-Link data record. Furthermore, identifying typical activity start times and matching those with working hours in Singapore can

further improve the accuracy for the detection of work activities.

Another important and inevitable indicator for activity type is the regularity and frequency of trips from and to the certain activity area. In order to be able to conduct such analysis, a longer term EZ-Link record of at least several consecutive days is required. This would enable the analysis of multiday travel behaviour and identification of the primary activity spaces or so called anchor points. The regularity of activities performed there would give a strong indication for their type. Furthermore multiday data would also help to identify shift workers with overnight shifts. Additionally, analysis of activities based on demographic factors and differentiation of work activities for adults and educational activities for students can be performed. Based on this analysis, the accuracy and reliability of estimations for home and work locations and their densities can be highly improved.

6. CONCLUSION AND FUTURE WORK

In this paper approaches for usage of public transport smart card fare payment data for characterization and analysis of travel behaviour and public transport system on example of city-state Singapore were presented. It was also demonstrated how data from travel surveys can be used to supplement the smart card data from AFC system and provide additional information for its interpretation. Furthermore a link between public transport usage and activities as reason for travelling was established and methods for identification of primary activity locations were presented. This is a first step towards an activity location model based on the public journeys record for entire Singapore. Disaggregated, multi-day data record, which is in prospect of becoming accessible for our research, will allow expansion and implementation of concepts presented in this work. Furthermore, obtained models of work locations on basis of EZ-Link data can be compared with existing models used by land planners. Disaggregated multi-day data will open the door for deeper analysis of travel patterns and travel behaviour, identification of load profiles for bus routes, stations and interchanges as well as allow development of route choice model for public transport. This will add additional value to EZ-Link data records and contribute to development of methods for processing and analysis of smart card data records for transportation planning purposes.

7. ACKNOWLEDGEMENT

The authors thank Land Transportation Authority of Singapore for supporting this research and providing the data set. Funding for this research was provided by National Research Foundation of Singapore.

REFERENCES

- Ahas, R. and Silm, S. and Järv, O. and Saluveer, E. and Tiru, M. (2010) Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones. *Journal of Urban Technology* 17, pp. 3-27.
- Bagchi, M. and White, P.R. (2005) The potential of public smart card data. *Transport Policy* 12, pp.464-474.
- Chu, K.K.A. and Chapleau, R. (2010). Augmenting Transit Trip Characterization and Travel Behavior Comprehension. *Journal of Transportation Research Board* 2183, pp. 29-40.
- Jang, W. (2010) Travel Time and Transfer Analysis Using Transit Smart Card Data. *Journal of Transportation Research Board* 2144, pp. 142-149.

- Li, D. and Lin, Y. and Zhao, X. and Song, H. and Zou, N. (2011) Estimating a transit passenger trip origin-destination matrix using automatic fare collection system. *Proceedings of the 16th international conference on Database systems for advanced applications*, pp. 502-513.
- Land Transportation Authority (2009) Household Interview Travel Survey 2008. Singapore.
- LTA Academy Singapore (2011) Singapore Land Transport Statistics as at the end 2010. *JOURNEYS* 6, pp.64-65.
- MATSim-T (2011) Multi Agent Transportation Simulation Toolkit, webpage, <http://www.matsim.org>.
- Munizaga, M. and Palma, C. and Fischer, D. (2011) Estimation of a Disaggregate Multimodal Public Transport OD Matrix from Passive Smart Card Data from Santiago, Chile. *Transportation Research Board Annual Meeting 2011* 11-0430.
- Páez, A. and and Trépanier, M. and Morency, C. (2011) Geodemographic analysis and the identification of potential business partnerships enabled by transit smart cards. *Transportation Research Part A: Policy and Practice* 45, pp. 640-652.
- Park, J.-Y. and Kim, D.-J. and Lim, Y. (2008) Use of Smart Card Data to Define Public Transit Use in Seoul, South Korea. *Journal of Transportation Research Board* 2063, pp. 3-9.
- Pelletier, M.-P. and Trépanier, M. and Morency, C. (2011) Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies* 19, pp. 557-568.
- Prakasam, S. (2008) The Evolution of e-payments in Public Transport – Singapore’s Experience. *Japan Railway & Transport Review* 50, pp. 36-39.
- Public Transport Council, Singapore (2010) , webpage, http://www.ptc.gov.sg/_files/Fare_revision_2010_news_release.pdf, accessed on 19.07.2011.